

A SYSTEM AND METHOD FOR AUDIO/VIDEO SPEAKER DETECTION

5

ABSTRACT OF THE DISCLOSURE

A system and method for detecting speech utilizing audio and video inputs. In one aspect, the invention collects audio data generated from a microphone device. In another aspect, the invention collects video data and processes the data to determine a mouth location for a given speaker. The audio and video are inputted into a time-delay neural network that processes the data to determine which target is speaking. The neural network processing is based upon a correlation to detected mouth movement from the video data and audio sounds detected by the microphone.